Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)
Volume 03, No. 3 (2014), hal 169 – 174.

METODE *PARTIAL LEAST SQUARES* UNTUK MENGATASI MULTIKOLINEARITAS PADA MODEL REGRESI LINEAR BERGANDA

Romika Indahwati, Dadan Kusnandar, Evy Sulistianingsih

INTISARI

Multikolinearitas merupakan salah satu permasalahan dalam analisis Regresi Linear. Multikolinearitas dapat menyebabkan estimasi parameter dengan metode Ordinary Least Squares (OLS) menjadi penduga yang masih tetap tak bias dan konsisten, tetapi tidak efisien. Salah satu metode yang dapat digunakan untuk mengatasi multikolinearitas, yaitu Metode Partial Least Squares (PLS). Pada penelitian ini tingkat efisiensi metode OLS dan PLS dibandingkan dalam mengestimasi parameter regresi ketika terdapat multikolinearitas dalam data. Penelitian ini menggunakan 21 kondisi data yang berbeda dalam ukuran sampel dan tingkat korelasi. Tingkat efisiensi dari kedua metode dibandingkan berdasarkan nilai bias dan Mean Square Error (MSE) dari nilai estimasi yang dihasilkan. Penelitian ini menunjukkan bahwa metode OLS merupakan penduga yang efisien ketika tidak ada korelasi antar variabelnya. Selain itu dapat disimpulkan juga bahwa metode PLS memiliki nilai bias yang cenderung mengecil seiring bertambahnya jumlah sampel dan sebaiknya digunakan sebagai suatu metode analisis ketika variabel bebas berkorelasi lebih dari atau sama dengan 0,8.

Kata Kunci*: Analisis Regresi, Matriks Korelasi, Multivariat, Simulasi*

PENDAHULUAN

Analisis regresi merupakan metode statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara suatu variabel dengan variabel lainnya. Variabel-variabel regresi yang berhubungan secara linear disebut sebagai regresi linear. Regresi linear yang menghubungkan satu variabel terikat dengan satu variabel bebas disebut regresi linear sederhana, sedangkan regresi linear yang menghubungkan satu variabel terikat dengan dua atau lebih variabel bebas disebut regresi linear berganda. Salah satu metode yang digunakan untuk mengestimasi parameter regresi adalah *Ordinary Least Squares* (OLS) [1].

Salah satu permasalahan yang perlu mendapatkan perhatian khusus dalam analisis regresi linear berganda adalah ketika ada multikolinearitas dalam variabel bebas. Keadaan ini biasanya terjadi ketika dalam model regresi yang digunakan terdapat suatu variabel bebas yang berkorelasi sangat tinggi dengan variabel bebas lainnya. Secara ekstrim, multikolinearitas antar variabel bebas dapat mengakibatkan pengaruh dari masing-masing variabel bebas terhadap variabel terikatnya menjadi sulit untuk dibedakan [2].

Ada beberapa metode untuk mengatasi masalah multikolinearitas, salah satunya adalah metode *Partial Least Squares* (PLS). Metode PLS merupakan metode yang mengkombinasikan sifat-sifat dari *Principal Component Analysis* (PCA) dan regresi linear berganda. Tujuan dari metode PLS adalah mengestimasi dan menganalisis variabel terikat dari variabel-variabel bebas. Dalam hal ini, PLS mereduksi dimensi variabel-variabel bebas dengan membentuk variabel-variabel baru yang merupakan kombinasi linear dari variabel-variabel bebas dengan dimensi lebih kecil, kemudian menggunakan metode OLS dalam mengestimasi variabel baru tersebut. [3]

Yeniay dan Atilla (2002) telah melakukan penelitian tentang perbandingan metode *Partial Least Squares*, *Principal Component Regression* (PCR) dan *Ridge Regression* (RR). Data yang digunakan adalah data real sebanyak 80 pengamatan dari *Gross Domestic Product Per Capita* (GDPPC) untuk setiap provinsi di Turkey. Hasil penelitian tersebut menyimpulkan bahwa metode PLS memiliki nilai koefisien determinasi yang tinggi, serta *Mean Square Error Prediction* (MSEP) dan *Root Mean Square Error Prediction* (RMSEP) yang lebih kecil dibandingkan dengan metode PCR dan RR [4].

Penelitian ini bertujuan untuk mengkaji metode PLS dalam mengatasi multikolinearitas pada regresi linear berganda serta membandingkan tingkat efisiensi dari metode OLS dan PLS melalui nilai bias dan MSE dari hasil estimasi. Pada penelitian ini, data yang digunakan merupakan data hasil simulasi yang dibangkitkan dengan menggunakan program statistik R versi 3.1.0. Simulasi dalam penelitian ini menggunakan ukuran pengamatan sebanyak 30, 100 dan 200 pengamatan. Variabel yang digunakan terdiri dari tiga variabel bebas x dan satu variabel terikat y, variabel bebas tersebut dibuat saling berkorelasi dengan koefisien korelasi yang digunakan meliputi $\pm 0{,}9$, $\pm 0{,}6$, $\pm 0{,}3$ dan 0. Parameter yang digunakan dalam simulasi adalah $\beta_0 = 0$, $\beta_1 = 1, \beta_2 = 1$ dan $\beta_3 = 1$ dengan banyaknya replikasi yang digunakan adalah sebanyak 10.000 kali.

MODEL REGRESI LINEAR BERGANDA

Regresi linear berganda merupakan regresi linear yang terdiri dari satu variabel terikat dan lebih dari satu variabel bebas. Variabel bebas dinotasikan dengan x dan variabel terikat dinotasikan dengan y. Secara umum, model regresi linear berganda yang melibatkan sejumlah $(k-1)$ variabel bebas adalah sebagai berikut [1]:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{k-1} x_{ik-1} + e_i, \tag{1}$$

dengan $i = 1,2, \ldots, n$, dimana n adalah banyaknya pengamatan; $\beta_0, \beta_1, \ldots, \beta_{k-1}$ merupakan parameter yang nilainya tidak diketahui dan e adalah nilai variabel acak yang merepresentasikan faktor-faktor lain yang mempengaruhi nilai variabel terikat dan disebut sebagai *residual*. Persamaan (1) dapat ditulis dalam notasi matriks sebagai berikut [2]:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}, \tag{2}$$

dengan $\boldsymbol{y}$ adalah vektor variabel terikat berukuran $(n \times 1)$, $\boldsymbol{X}$ adalah matriks variabel bebas yang berukuran $(n \times k)$ dimana setiap kolomnya merupakan nilai-nilai pengamatan bagi masing-masing variabel x, kecuali kolom pertama dari matriks $\boldsymbol{X}$ yang merupakan kolom bernilai satu. Sedangkan $\boldsymbol{\beta}$ merupakan vektor parameter berukuran $(k \times 1)$ dan $\boldsymbol{e}$ adalah vektor *residual* berukuran $(n \times 1)$.

Salah satu metode yang sering digunakan untuk mengestimasi parameter regresi adalah metode OLS, dengan model persamaan penduga regresi linear berganda adalah sebagai berikut:

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}, \tag{3}$$

dimana $\widehat{\boldsymbol{y}}$ adalah vektor nilai estimasi pengamatan dari $\boldsymbol{y}$ dan $\widehat{\boldsymbol{\beta}}$ adalah vektor penduga dari $\boldsymbol{\beta}$. Prinsip dasar metode OLS adalah meminimumkan jumlah kuadrat *residual*. Metode ini mengestimasi vektor parameter $\boldsymbol{\beta}$ dengan persamaan sebagai berikut [1]:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{OLS}} = (\boldsymbol{X}'\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}'\boldsymbol{y} \tag{4}$$

METODE *PARTIAL LEAST SQUARES* (PLS)

Metode *Partial Least Squares* (PLS) pertama kali diperkenalkan oleh Herman Ole Andreas Wold pada tahun 1960 sebagai metode alternatif untuk mengatasi keterbatasan metode *Ordinary Least Squares* (OLS) ketika data mengalami masalah multikolinearitas. Untuk meregresikan variabel terikat y dengan variabel bebas $x_1, x_2, \cdots, x_k$, metode PLS mencari komponen-komponen baru yang berperan sebagai variabel bebas untuk mengestimasi parameter regresi [4].

Jika terdapat sejumlah k variabel bebas dan sebuah variabel terikat, dalam prosesnya metode PLS mengasumsikan semua variabel telah distandarisasi dalam bentuk sebagai berikut [5]:

$$y_i - \bar{y} = \beta_1^*(x_{i1} - \bar{x}_1) + \beta_2^*(x_{i2} - \bar{x}_2) + \cdots + \beta_k^*(x_{ik} - \bar{x}_k) + e_i^*, \tag{5}$$

dengan $i = 1,2, \ldots, n$. Persamaan (5) dapat disajikan dalam bentuk matriks sebagai berikut:

$$\boldsymbol{y}^* = \boldsymbol{X}^*\boldsymbol{\beta}^* + \boldsymbol{e}^*, \tag{6}$$

dengan $\boldsymbol{y}^*$ adalah vektor variabel terikat yang sudah terstandarisasi berukuran $(n \times 1)$, $\boldsymbol{X}^*$ adalah matriks variabel bebas yang sudah terstandarisasi berukuran $(n \times k)$, sedangkan $\boldsymbol{\beta}^*$ merupakan vektor parameter berukuran $(k \times 1)$ dan $\boldsymbol{e}^*$ adalah vektor *residual* berukuran $(n \times 1)$.

Sebelum menentukan penduga parameter $\boldsymbol{\beta}^*$ pada persamaan (6) terlebih dahulu dibentuk komponen utama yang akan digunakan sebagai variabel baru untuk mengestimasi parameter regresi $\boldsymbol{y}^*$ dengan menggunakan algoritma PLS1. Model regresi PLS dengan k komponen utama dapat dirumuskan sebagai berikut [6]:

$$\boldsymbol{y}^* = \boldsymbol{X}^*\boldsymbol{W}\boldsymbol{P}'\boldsymbol{\beta}^* + \boldsymbol{e}^* \tag{7}$$

dimana $\boldsymbol{W}$ merupakan matriks pembobot untuk matriks $\boldsymbol{X}^*$ yang berukuran $(k \times k)$ dan $\boldsymbol{P}'$ merupakan matriks muatan yang berukuran $(k \times k)$. Dengan mendefinisikan $\boldsymbol{X}^*\boldsymbol{W} = \boldsymbol{T}$ dan $\boldsymbol{P}'\boldsymbol{\beta}^* = \boldsymbol{c}$, maka persamaan (7) menjadi [6]:

$$\boldsymbol{y}^* = \boldsymbol{T}\boldsymbol{c} + \boldsymbol{e}^* \tag{8}$$

dengan $\boldsymbol{y}^*$ merupakan vektor variabel terikat berukuran $(n \times 1)$, $\boldsymbol{T}$ adalah mariks komponen utama berukuran $(n \times k)$ dan $\boldsymbol{c}$ adalah vektor koefisien regresi berukuran $(k \times 1)$.

Secara umum langkah-langkah pembentukan komponen-komponen utama dengan menggunakan algoritma PLS1 adalah sebagai berikut [7]:

1. Tentukan bobot $\boldsymbol{w}_j = \frac{\boldsymbol{X}^{*\prime}_{(j)}\boldsymbol{y}^*}{\left\|\boldsymbol{X}^{*\prime}_{(j)}\boldsymbol{y}^*\right\|}$.
2. Tentukan komponen $\boldsymbol{t}_j = \boldsymbol{X}^*_{(j)}\boldsymbol{w}_j$.
3. Tentukan koefisien regresi $\hat{\boldsymbol{c}}_j = \frac{\boldsymbol{t}_j'\boldsymbol{y}^*}{\boldsymbol{t}_j'\boldsymbol{t}_j}$.
4. Tentukan muatan $\boldsymbol{p}_j = \frac{\boldsymbol{X}^{*\prime}_{(j)}\boldsymbol{t}_j}{\boldsymbol{t}_j'\boldsymbol{t}_j}$.
5. Tentukan $\boldsymbol{X}^*_{(j+1)} = \boldsymbol{X}^*_{(j)} - \boldsymbol{t}_j\boldsymbol{p}_j'$.
6. Ulangi langkah pertama sampai langkah kelima untuk $j = 1,2,\cdots,k$. Iterasi berhenti ketika $j = k$.

Hasil keseluruhan iterasi dari algoritma PLS1 menghasilkan suatu matriks $\boldsymbol{W}$ dan $\boldsymbol{P}$ yang berukuran $(k \times k)$ dan matriks $\boldsymbol{T}$ yang berukuran $(n \times k)$ serta vektor kolom $\boldsymbol{c}$ berukuran $(k \times 1)$. Matriks $\boldsymbol{W}, \boldsymbol{P}$ dan $\boldsymbol{T}$ serta vektor $\boldsymbol{c}$ tersebut akan membentuk beberapa model regresi. Jika model regresi yang digunakan terdiri dari tiga variabel bebas, maka model regresi PLS yang diperoleh adalah sebagai berikut:

1) Model regresi PLS dengan 1 komponen: $\boldsymbol{y}^* = \boldsymbol{t}_1\boldsymbol{c}_1 + \boldsymbol{e}^*$
2) Model regresi PLS dengan 2 komponen: $\boldsymbol{y}^* = \boldsymbol{t}_1\boldsymbol{c}_1 + \boldsymbol{t}_2\boldsymbol{c}_2 + \boldsymbol{e}^*$
3) Model regresi PLS dengan 3 komponen: $\boldsymbol{y}^* = \boldsymbol{t}_1\boldsymbol{c}_1 + \boldsymbol{t}_2\boldsymbol{c}_2 + \boldsymbol{t}_3\boldsymbol{c}_3 + \boldsymbol{e}^*$

Selanjutnya dipilih salah satu model terbaik dengan melihat nilai proporsi variansnya. Jika nilai proporsi varians yang dihasilkan pada model regresi PLS dengan satu komponen lebih besar dari sama dengan 80%, maka dengan hanya menggunakan vektor komponen pertama $(\boldsymbol{t}_1)$ sudah cukup untuk menjelaskan variasi $\boldsymbol{X}^*_{(1)}$. Sebaliknya, jika proporsi yang diperoleh kurang dari 80% maka dibutuhkan model regresi PLS dengan dua komponen [7].

Setelah model regresi PLS diperoleh, selanjutnya dilakukan estimasi parameter regresi dengan model penduganya sebagai berikut:

$$\widehat{\boldsymbol{y}}^* = \boldsymbol{T}\hat{\boldsymbol{c}} \tag{9}$$

sehingga diperoleh penduga parameter $\boldsymbol{c}$, yaitu:

$$\hat{\boldsymbol{c}} = (\boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{T}'\boldsymbol{y}^*. \tag{10}$$

Berdasarkan definisi $\boldsymbol{\beta}^*\boldsymbol{P}' = \boldsymbol{c}$, dimana $(\boldsymbol{P}')^{-1} = \boldsymbol{W}$, diperoleh penduga $\widehat{\boldsymbol{\beta}}^*$ dari regresi PLS adalah sebagai berikut:

$$\widehat{\boldsymbol{\beta}}^* = \boldsymbol{W}\hat{\boldsymbol{c}}. \tag{11}$$

Selanjutnya dengan mensubstitusikan persamaan (10) ke persamaan (11) diperoleh penduga $\widehat{\boldsymbol{\beta}}_{\boldsymbol{PLS}}$ yaitu:

$$\widehat{\boldsymbol{\beta}}^* = \boldsymbol{W}(\boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{T}'\boldsymbol{y}^* \tag{12}$$

SIMULASI DATA

Pada penelitian ini, simulasi dilakukan mengikuti model regresi linear berganda yang melibatkan hanya tiga variabel bebas untuk setiap kondisi data dengan proses simulasi dan estimasi secara umum dilakukan sebagai berikut:

1. Ditetapkan nilai parameter regresi yang digunakan yaitu $\beta_0 = 0$, dan $\beta_1 = \beta_2 = \beta_3 = 1$ untuk setiap simulasi.
2. Variabel bebas x_1, x_2, x_3 dibangkitkan mengikuti distribusi Normal Multivariat untuk setiap sampel (n) yaitu: 30, 100 dan 200, dengan vektor rata-rata (μ) dan matriks varians kovariansnya $(\sum)$ dirancang agar memiliki nilai koefisien korelasi antar dua variabel, dengan tingkat korelasi yang digunakan adalah: $(\rho_i) = \pm 0{,}9,\ \pm 0{,}6,\ \pm 0{,}3$ dan 0. Untuk vektor rata-rata dan matriks varians kovariansnya adalah seperti berikut:
$$\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \sum = \begin{bmatrix} 1 & \rho_i & \rho_i \\ \rho_i & 1 & \rho_i \\ \rho_i & \rho_i & 1 \end{bmatrix}$$
3. Bangkitkan *residual* (e) mengikuti distribusi Normal Standar $e_i \sim N(0{,}1)$ dengan rata-ratanya $(\mu) = 0$ dan variansnya $(\sigma^2) = 1$.
4. Variabel y diperoleh dengan mensubstitusikan nilai-nilai dari variabel bebas x dan *residual* (e) ke dalam model berikut:
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e$$
5. Mengestimasi parameter model regresi dengan metode *Ordinary Least Squares* (OLS) dan *Partial Least Squares* (PLS) untuk setiap kondisi data.
6. Menghitung nilai bias dan MSE untuk metode OLS dan PLS.
7. Untuk setiap kombinasi dari n dan ρ, dilakukan perulangan sebanyak 10.000 kali pengulangan.

HASIL SIMULASI

Tabel 1 menyajikakan nilai bias parameter penduga metode OLS dan PLS. Tabel 1 menunjukkan bahwa nilai bias untuk metode OLS dan PLS pada koefisien korelasi (ρ) dan ukuran sampel (n) yang berbeda-beda. Dapat dilihat bahwa mutlak dari nilai bias metode PLS lebih besar dibandingkan metode OLS ketika koefisien korelasi bernilai $-0{,}9, -0{,}6, -0{,}3,$ 0, dan 0,3 untuk setiap ukuran sampel. Ketika koefisien korelasi 0,6 nilai bias β_1 dari metode PLS lebih kecil dibandingkan dengan nilai bias metode OLS untuk setiap n, nilai bias β_3 dari metode PLS lebih kecil dibandingkan dengan nilai bias metode OLS dengan jumlah sampel 30 dan 100 , sedangkan β_2 untuk setiap n dan β_3 dengan jumlah sampel 200 nilai bias metode OLS lebih kecil dari pada nilai bias metode PLS. Pada saat koefisien korelasi 0,9 nilai bias dari metode PLS jauh lebih kecil dibandingkan dengan nilai bias dari metode OLS untuk setiap ukuran sampel.

Berdasarkan Tabel 1, diketahui bahwa metode OLS merupakan penduga yang lebih baik pada saat tidak terdapat korelasi antar variabel bebasnya. Hasil simulasi juga menunjukkan bahwa metode OLS jauh lebih baik dari metode PLS ketika variabel bebas berkorelasi negatif. Selain itu Tabel 1 juga menunjukkan nilai bias metode OLS secara keseluruhan lebih kecil dari 10%, yang berarti nilai bias tersebut masih dapat diterima. Hal tersebut membuktikan bahwa adanya multikolinearitas tidak mengganggu ketakbiasan penduga yang dihasilkan oleh metode OLS. Metode PLS memiliki nilai bias yang lebih kecil ketika terdapat koefisien korelasi positif yang tinggi antar variabel bebasnya. Dengan kata lain metode PLS lebih baik digunakan untuk mengestimasi parameter regresi ketika koefisien korelasi antar variabel bebasnya lebih besar dari sama dengan 0,8. Tabel 1 juga menunjukkan bahwa semakin meningkatnya ukuran sampel yang digunakan, semakin kecil nilai bias yang dihasilkan oleh metode PLS. Dengan kata lain, nilai bias penduganya cenderung mengecil seiring dengan bertambahnya ukuran sampel untuk setiap ukuran koefisien korelasi.

Tabel 1. Nilai Bias Parameter Penduga dengan Metode OLS dan PLS

ρ	n	β_1		β_2		β_3	
		OLS	PLS	OLS	PLS	OLS	PLS
-0,9	30	-0,00032	0,00548	0,00942	-0,12219	-0,00752	-0,13621
	100	0,00089	0,02916	0,00520	-0,08819	-0,00256	-0,10063
	200	-0,00305	0,03369	-0,00146	-0,08656	-0,00202	-0,08662
-0,6	30	-0,00017	-0,01242	-0,00135	-0,32442	0,00083	-0,32556
	100	-0,00079	0,03236	-0,00272	-0,33073	0,00208	-0,32688
	200	0,00001	0,05039	0,00035	-0,32679	-0,00086	-0,32894
-0,3	30	0,00088	-0,62911	0,00177	-0,15024	-0,00153	-0,15085
	100	-0,00097	-0,65394	0,00231	-0,10989	-0,00020	-0,11222
	200	-0,00022	-0,65828	-0,00157	-0,10110	-0,00034	-0,10130
0	30	-0,00197	-0,05636	0,00298	-0,05078	-0,00216	-0,05525
	100	-0,00075	-0,02851	0,00173	-0,02558	0,00049	-0,02719
	200	-0,00104	-0,01951	0,00042	-0,01816	-0,00138	-0,02025
0,3	30	-0,00065	-0,01836	-0,00249	-0,01911	0,00166	-0,01559
	100	0,00083	-0,00679	0,00019	-0,00719	-0,00121	-0,00855
	200	-0,00067	-0,00530	-0,00078	-0,00555	-0,00112	-0,00611
0,6	30	-0,00816	0,00121	0,00279	-0,00638	-0,01514	-0,00597
	100	-0,00487	-0,00093	-0,00281	-0,00691	0,00373	-0,00054
	200	-0,00305	-0,00048	0,00091	-0,00166	-0,00081	-0,00354
0,9	30	0,00734	0,00057	-0,00299	-0,00113	-0,00286	-0,00139
	100	0,00550	0,00019	0,00142	-0,00055	-0,00632	-0,00047
	200	-0,00615	-0,00019	0,00479	-0,00021	0,00133	-0,00029

Selain dari nilai bias, tingkat efisien metode OLS dan PLS juga dapat dilihat dari nilai MSE sebagaimana yang terlihat pada Tabel 2 berikut:

Tabel 2. Nilai MSE Parameter Penduga dengan Metode OLS dan PLS

ρ	n	β_1		β_2		β_3	
		OLS	PLS	OLS	PLS	OLS	PLS
-0,9	30	0,38530	0,40580	0,37915	0,33859	0,38942	0,35134
	100	0,16104	0,18341	0,16005	0,15770	0,15821	0,15817
	200	0,09995	0,11601	0,10273	0,10714	0,10087	0,10639
-0,6	30	0,07799	0,07798	0,07639	0,17765	0,07725	0,18081
	100	0,03267	0,03348	0,03298	0,14481	0,03287	0,14316
	200	0,02025	0,02221	0,02061	0,13088	0,02065	0,13158
-0,3	30	0,04727	0,53555	0,04882	0,08777	0,04749	0,08634
	100	0,02011	0,50269	0,02018	0,04294	0,01996	0,04425
	200	0,01254	0,48293	0,01276	0,03108	0,01246	0,03093
0	30	0,03967	0,04280	0,03880	0,04125	0,03974	0,04301
	100	0,01662	0,01765	0,01636	0,01719	0,01666	0,01758
	200	0,01022	0,01066	0,01061	0,01102	0,01043	0,01098
0,3	30	0,03967	0,04280	0,03880	0,04125	0,03974	0,04301
	100	0,01994	0,01942	0,01976	0,01930	0,01979	0,01931
	200	0,01244	0,01224	0,01298	0,01275	0,01297	0,01278
0,6	30	0,18426	0,04369	0,17954	0,04479	0,18417	0,04483
	100	0,07562	0,03277	0,07449	0,03290	0,07877	0,03302
	200	0,04781	0,02104	0,04724	0,02121	0,04809	0,02091
0,9	30	0,54558	0,00573	0,56058	0,00573	0,55633	0,00559
	100	0,22427	0,00242	0,22711	0,00246	0,23018	0,00246
	200	0,14573	0,00156	0,13898	0,00157	0,14130	0,00158

Tabel 2 menunjukkan nilai MSE untuk metode OLS dan PLS pada koefisien korelasi dan ukuran sampel yang berbeda-beda. Tabel 2 menunjukkan bahwa semakin besar ukuran sampel, semakin kecil nilai MSE yang dihasilkan. Namun metode PLS memiliki nilai MSE yang lebih kecil dibandingkan metode OLS untuk data dengan koefisien korelasi yang positif. Ketika tidak ada koefisien korelasi antar variabel bebasnya metode OLS merupakan penduga yang efisien karena memiliki nilai MSE yang kecil dibandingkan metode PLS. Selain itu metode OLS juga memiliki nilai MSE yang lebih kecil dibandingkan metode PLS ketika terdapat koefisien korelasi negatif antar variabel bebasnya. Ketika koefisien korelasi $0{,}3$ nilai MSE dari metode PLS bernilai lebih kecil dari pada nilai MSE dari metode OLS untuk setiap ukuran sampel. Begitu pula untuk koefisien korelasi $0{,}6$ dan $0{,}9$ nilai MSE dari metode PLS bernilai lebih kecil dibandingkan metode OLS. Dengan kata lain metode PLS memiliki tingkat kesalahan yang kecil untuk mengestimasi parameter regresi ketika koefisien korelasinya $(\rho) \geq 0{,}3$.

PENUTUP

Berdasarkan hasil analisis dan pembahasan dalam penelitian ini dapat ditarik kesimpulan bahwa penduga parameter regresi yang dihasilkan oleh metode PLS menjadi bias dan tidak efisien ketika digunakan untuk mengestimasi parameter regresi ketika terdapat korelasi negatif, tidak ada korelasi dan korelasi kecil antar variabel bebasnya. Hal tersebut terlihat dari nilai bias dan MSE yang dihasilkan metode PLS lebih besar dibandingkan dengan metode OLS

Suatu penduga dikatakan baik jika estimasinya menghasilkan nilai bias dan MSE yang kecil atau mendekati nilai nol. Metode PLS merupakan metode penduga yang baik ketika terdapat koefisien korelasi lebih besar dari atau sama dengan $0{,}8$ antar variabel bebasnya. Hal ini diketahui dari nilai bias dan MSE yang kecil dan mendekati nol.

DAFTAR PUSTAKA

[1]. Kutner, MH. Nachtsheim, CJ. Neter, J. dan Li, W. *Applied Linear Regression Models*. Newyork: McGraw-Hill Companies. Inc; 2004.

[2]. Kusnandar, D. *Metode Statistik dan Aplikasinya dengan Minitab dan Excel*. Madyan Press, Yogyakarta; 2004.

[3]. Abdi, H. *Partial Least Squares Regression (PLS Regression)*. Encyclopedia of Measurement and Statistics. Thousand Oaks (CA): Sage; 2007.

[4]. Yeniay, O. dan Attila, G. A Comparison of Partial Least Squares Regressions with Other Prediction Methods. *Hacettepe Journal of Mathematics and Statistics*. 2002; 31: 99-111.

[5]. Echambadi, R. dan Hess, JD. Mean-Centering Does Not Alleviate Collinearity Problems in Moderated Multiple Regression Models, *Marketing Science*. 2007; 26: 438-445.

[6]. Hoskuldsson, A Partial Least Squares Regression Methods, *Journal of Chemometrics*. 1988; 2: 211-228.

[7]. Varmuza, K. dan Filzmoser, P. *Introduction To Multivariate Statistical Analysis In Chemometrics*, Taylor & Francis Group, Boca Raton London New York; 2008.

ROMIKA INDAHWATI : Jurusan Matematika, FMIPA UNTAN, Pontianak, mikha.1005@gmail.com

DADAN KUSNANDAR : Jurusan Matematika, FMIPA UNTAN, Pontianak, dkusnand@yahoo.com

EVY SULISTIANINGSIH : Jurusan Matematika, FMIPA UNTAN, Pontianak, evysulistianingsih@gmail.com